

REVIEW DRAFT

To: PATH steering committee
From: Charlie Paulsen, Al Giorgi, Jim Anderson
Subj: An empirical approach to identifying and weighting critical uncertainties
Date: Monday, April 20, 1998
File Name: **wt042098.doc**

In reviewing the latest (2/8/98) results, it seems to us that trying to weight all 5,148 aggregate hypotheses is at best a Herculean task. In practice, it is likely to be a job that is beyond the resources of the PATH group in any reasonable time frame. It also appears unlikely that we will ever arrive at a strong consensus on an appropriate set of weights, given the widely divergent views of different PATH members on reasons for the decline of Snake River stocks. In addition, the Implementation Team's decision that we should devote most of our attention to fall chinook suggests that we use available spring chinook time as efficiently as possible. Finally, we believe that there is empirical information that has not yet been brought to bear on the weighting of different hypotheses.

Therefore, we have developed ways to bound the problem, so that we can analyze a manageable subset of the 5K+ hypotheses. We have also developed methods for quantitatively analyzing the relative degree of support for both simple and aggregate hypotheses. The intent of these efforts is not to replace the weight-of-evidence approach, but to bound the problem and supplement it with additional empirical data and statistical inferences. This memo suggests how we could proceed. We welcome comments on the approaches. We divided the problem into five parts.. Each part is explained more fully in the corresponding section of the memo.

1. What uncertainties are "critical," in the sense that they distinguish prospective scenarios with high probabilities of meeting survival and recovery goals from those with lower probabilities? Are some critical uncertainties more important than others? This memo includes an initial examination of the problem of finding critical uncertainties. A memo by Hinrichsen and Paulsen (forthcoming) explores the issue in more detail, including methods to *objectively* analyze the importance of different hypotheses. Note that this issue is distinct from the problem of using empirical data in the weighting process.

2. Each of the 5K+ prospective aggregate hypotheses has a corresponding retrospective aggregate hypothesis. Some cannot be tested with retrospective data (e.g., duration of pre-removal period). We ask, among those that are critical [from (1), above], and can be tested with data used to fit the retrospective models, which fit the data better? (i.e., which have superior AICs, BICs, etc.?) This question involves both passage model fits to past data and fits of BSM (Bayesian Simulation Model) parameters (e.g., λ_n 's). It is inevitable that some aggregate hypotheses will comport better than others with retrospective data. This may provide additional information to assign weights to particular hypotheses.
3. The passage models were calibrated using data through passage years 1995-96. How well do CRiSP and FLUSH perform in predicting data not used in model validation and calibration, such as 1996-98 in-river survivals, PIT tag recovery proportions from 1989 on, TCRs for passage years 1994-95, etc? This information could be used to differentially weight hypotheses.
4. The BSM was calibrated using data through brood year 1990. To initialize prospective model runs, it used estimated spawners from 1991-95. The aggregate spring/summer chinook run in 1996 (6K hatchery + wild) was only 1/7th of that in 1997 (42k+). Presumably, similar variation exists in recruitment and spawning escapement for the seven Snake index stocks in those years. How well does the BSM perform in predicting that recruitment and spawning escapement when using the relevant, important aggregate hypotheses identified in (1) and (2)?
5. We have stock-recruit (S/R) data for three mid-Columbia spring chinook stocks, and have developed Delta and Alpha life-cycle models for them. This information has not been analyzed by PATH as intensively as the Snake stocks, however. Initial examination of the S/R data suggests that these stocks, while broadly similar to the Snake index stocks, did not decline as quickly or as dramatically as dams went in on the Columbia above McNary. Furthermore, it appears that many of the relevant Snake aggregate hypotheses should apply equally well to the mid-Columbia stocks. We suspect that a closer examination of this information might lend support to some aggregate hypotheses for the Snake, while others are probably not supported to the same degree. Is this really the case?

These analyses will be useful to the extent that we can use them to narrow the range of uncertainties in BSM projections of the outcomes of management actions. At present, it is difficult to use the prospective analysis to select a hydro management action. Unless we can confidently assign different weights to the different hypotheses, it will continue to be hard to make a recommendation based on the model results. In reviewing the remainder of the document, this is important to keep that in mind. The purpose of this is not to find a single, “best” aggregate hypothesis, but to narrow the range to a manageable number, and to assess the relative support for different hypotheses based on measured data and simple statistical analyses. It may even be possible for PATH members to agree on some of the weights. The goal is to use the results to make more confident, scientifically sound recommendations on management actions.

1. Critical Uncertainties

Work done in Chapter 5 (February 1998 version) suggests that some uncertainties are substantially more important than others, especially when assessing whether or not a management action (A1, A2...) meets survival and recovery objectives. In particular, of the 15 or so hypotheses or uncertainties analyzed to date (from Table 4.1, leading to 4680 aggregate prospective hypotheses), few make a difference in meeting objectives. In particular, passage model (really a complex of in-river survival, TCR model, and “D” values), FGE, predator removal, prospective model, and extra mortality/future climate differentiate hypotheses meeting survival and recovery standards from those that do not¹.

We summarize this in **Table 1**. Of the 11 classes of uncertainties (combining extra mortality and climate), only five make a big difference in the combinations of actions and hypotheses that meet the survival and recovery standards. What we mean by this is that the distribution of uncertainties is quite different for the set of actions meeting the objectives, as opposed to the entire set of BSM outputs. The important uncertainties are the choice of “passage” model (obviously, a complex of other hypotheses), FGE, predator removal, prospective model (alpha and delta), and extra mortality/climate. The first four of these each have two discrete levels (e.g., CRiSP/FLUSH), while the last has

¹ These results may change as hypotheses are disaggregated and explored further. See Hinrichsen and Paulsen for details.

five. The result is 80 ($2^4 * 5$) aggregate hypotheses, times three management actions, rather than 5K+. Clearly, it will be easier to analyze 80 aggregate hypotheses than to consider 5,000 in detail. Note that management actions are not hypotheses, but human choices about what to do.

Obviously, there is some judgement involved in arriving at a conclusion on the number of uncertainties that are critical. For example, we view the habitat uncertainty as unimportant (at least as analyzed to date), although the runs meeting the standards are divided 53 vs. 47 percent, not 50/50. In addition, some of the differences (e.g., transport model) occur because CRiSP appears more often than FLUSH in the set that meets the goals. However one judges the importance of individual hypotheses, the method looks to be a promising way to bound the weighting exercise, especially when applied to each alternative action. (See Hinrichsen and Paulsen for additional analysis of the importance of individual hypotheses and questions concerning analysis of the existing BSM output.)

2. BSM fits to data through BY90 In Chapter 5 of the retrospective report, Rick displayed the fit of the delta model (and variants) to stock-recruit (S/R) data for a variety of model parameterizations. On that basis, he selected a modest number of promising models for more detailed analysis. Although the number of hypotheses and uncertainties has proliferated since then, we have not yet seen similar fits of S/R models – both Alpha and Delta – to the much larger number of aggregate hypotheses now included.

We propose to check these model fits (using data through brood year 1990) for the testable subset of the 80 aggregate hypotheses identified above. Note that management actions don't enter the calculations, and that to compare the BSM hypotheses with data it will be necessary to choose values for the other uncertainties. However, given the prospective results, we do not expect that these other uncertainties will make much difference (see **Table 2**). In addition, many of them do not apply to retrospective analyses. These are noted in column 2 of Table 2. For example, the “BKD” hypothesis essentially states that extra mortality will have the same value in future as it has had in the recent past. While this clearly is important for the success of future management actions, its predictions cannot be compared to retrospective data. In

addition, since ESBS bypass and predator removal did not come on-line until after brood year 1990 (passage year 1992), they cannot be compared to the data used to fit the BSM².

Therefore, among the important uncertainties, one can only test a modest subset: passage model (2), prospective model (2), and two of the extra mortality/climate hypotheses (hydro and regime shift). The result is that only eight of the retrospective models are readily "weighable"³. Nevertheless, the results from this exercise should be useful in weighting the many prospective models, since it is impossible that they will all fit the data equally well. More details are shown in **Table 3**, including model equations. As already noted, we do not expect these to be "definitive" in the sense that some hypotheses would be given a weight of zero, while others would be assigned a one. What we do expect is that we will be able to give non-uniform weights to some of the models and uncertainties, after looking at comparisons of $V(n)$'s, lambdas, AIC/BIC scores, etc.

In addition to analyzing the hypotheses as formulated in existing BSM runs, some additional sensitivities are possible as well. We mention two examples; others are obviously possible. First, a recent analysis by Hinrichsen ("Influence of exceptional spawner-recruit data for the John Day Middle Fork," 12/31/97) establishes that Delta model estimates of μ and nX would change substantially if several years of data (of questionable quality) for one stock were eliminated from the analysis. What would the effect of this be on the prospective model results? It is our understanding that some analysis of this has been performed, but we have not yet seen the results. A second possibility is related to the paucity of "out-of-sample" data for testing the models. In the course of developing retrospective models, we did not "reserve" any information for verifying model results. The question then arises, what if one were to estimate retrospective alpha and delta models withholding (say) the last five years of S/R data. How well would the models' predictions compare to what actually occurred? Alternatively, how well do the models' predictions compare to the S/R time series of the 8-10 Snake stocks where data were developed only last year? We believe that these could be used to test and weight the various prospective life cycle models.

² However, see section (3) for passage model testing.

³ Note that one could potentially compare the turbine (TURB1-TURB6) and transportation model (0, 1, and 3) hypotheses to S/R data, but these do not appear to be important in the prospective analysis.

3. **Passage model fits to data used for calibration and new information**

While the passage models have been calibrated to reach survival and TCR estimates through passage years 1995-96, there are a number of data series that have not, to date, been used to challenge the models systematically. These include the following:

- Recovery proportions of PIT-tagged fish, 1989-97, from various Snake and Columbia projects as far down the system as data permits;
- TCRs for transport study groups, passage year 1995 (should have complete recoveries of 3-ocean fish in 1998);
- Fish travel times for PIT-tagged fish, starting in 1987 (from Snake River traps);
- Reach survivals of PIT-tagged fish for 1996-98, as low in the system as data permits.
- Potentially, recoveries of fish below Bonneville in experimental trawls and by the Rice Island tern colony. Given the small number of recoveries to date, these would be pilot studies.

Obviously, there will be a number of decisions regarding time step, level of detail, etc. in making comparisons of model predictions to data. These would likely include:

- Time step to be used (seasonal average or adapted to PIT tag study groups); CRiSP will have an easier time adapting to many individual releases groups, while FLUSH will find it easier to deal with seasonal averages.
- Recoveries of tagged fish will be non-zero but “low” for many years and locations (e.g., lower river dams). We will need to explore whether estimates derived from these sites are precise enough to be useful in discriminating among hypotheses.
- Out-of-sample estimates are limited, especially in low-flow years. Are they enough to help weight hypotheses?
- Passage routes of tagged fish often differ from those of the run at large. Therefore, passage survival of PIT tagged fish is not, in general, equal to $V(n)$, survival of in-river migrants: the tagged fish are generally returned to the river after being bypassed, while the run at large was usually transported following bypass at Lower Granite and Little Goose. If $V(n)$ is a desired output from the comparisons, adjustments will be required to account for passage routes (turbine, spillway) of fish that were not bypassed (and hence transported).

The plan is to use the passage model versions that were employed to generate the February 1998 BSM results. However, we expect that the models may be revised after the comparisons are completed, and before doing further BSM runs.

4. BSM predictions of recent S/R data

The BSM has been calibrated for retrospective runs with data through brood year 1990, and is initialized with estimated spawning escapements (for prospective analyses) with data from run years 1991-95. Spawning escapement should be easy to calculate for 1996-97. Therefore, several “out of sample” tests are possible with existing information. First, it should be possible to predict BY91 and BY92 recruitment for the various retrospective models, and compare how well their respective predictions compare to estimated returns to the Columbia. In addition, recruitment of 1-ocean and 2-ocean fish (ages 3 and 4) is now complete for BY93. It should be easy to predict these numbers to see how they compare to current estimates. Finally, the same calculation should also yield predictions for 3-ocean fish from BY93, returning in 1998 (see **Table 4**). We can make those predictions now, and see how they compare to estimates from 1998 harvest, conversion, and spawning escapement numbers when these become available in the fall of 1998. The predictions of median recruitment would obviously be accompanied by variance estimates. These predictions of recruits and spawners could either be linked to passage model predictions of inriver survival and TCRs from (3), above, or could use the same scheme as is employed for long-term passage survival projections in the BSM.

It will obviously require some time to develop the S/R data and compile the predictions. Given the IT’s directive, we’ll need to arrange schedules so that this does not cut into fall chinook work. However, the different hypotheses (especially those associated with different passage and life cycle models) should provide different estimates of stock performance. A comparison of those estimates to actual stock performance will provide additional information to weight hypotheses and, ultimately, provide management recommendations for spring/summer chinook.

5. Mid-Columbia spring chinook analysis

Olaf Langness has developed spawning escapement and recruitment data for three mid-Columbia (or Upper Columbia) spring chinook stocks. While these stocks have

received a modest amount of analysis (i.e., delta model results for use in the habitat workgroup) they have not been analyzed at the same depth as the Snake stocks. Since age data are not available for individual stocks for all years, some caution is required in interpreting results from stock-recruit (S/R) models. Still, some intriguing patterns are apparent. (**Table 5 and Figure 1**). For example, although recruit/spawner (R/S) ratios were clearly much higher in the 60s and early 70s than in later years, the R/S ratios were well above one for most stocks into the mid-80s. In addition, spawning escapement did not decline markedly until the early 90s for the Wenatchee and Methow stocks. These patterns are quite different from the Snake, where declines began in the early 70s and where the drop in spawner numbers was much more abrupt. On the other hand, the post-90 drop in spawning escapement was severe for both stock groups.

Preliminary examination of the data and potential causes for the mid-Columbia decline suggest that most of the hypotheses developed for the Snake stocks should be applicable to the mid-Colombia: passage models, prospective models, extra mortality, etc. We suspect that applying the Snake retrospective models (or a subset thereof) to the mid-Columbia would provide a quasi-independent test of the hypotheses regarding the decline and possible futures for the Snake.

Conclusions

Preliminary analysis suggests that a small subset of the uncertainties are responsible for much of the difference in BSM projections. Of these, a subset can be compared to data that are either already available or that will be obtained from routine monitoring in 1998. We believe that this information can, and should, be used to inform PATH decisions on weighting individual hypotheses, and hence in weighting the results of the BSM projections. We hope that by doing this and by challenging the passage models, we will have additional information to present to decision-makers regarding the likely outcomes of hydro-related management strategies.

Table 1. Distribution of BSM runs for all models and results meeting survival and recovery standards. See text for details.

Uncertainty/Hypothesis*	Hypothesis	% of 5K+ model runs	% of runs meeting survival and recovery standards
Passage Model	PMOD1 – CRiSP	50	61
	PMOD2 – FLUSH	50	39
Fish Guidance Efficiency	ESBS > STS	50	59
	ESBS + STS	50	41
Turbine/Bypass Survival	TURB1	37.2	34
	TURB4	21.4	27
	TURB5	31.4	33
	TURB6	10	6
Predator Removal	PREM1 (no res. mortality reduction)	50	31
	PREM3 (25% reduction)	50	69
Pre-Removal Period	N/A (A1 and A2)	67	57
	PRER1	17	26
	PRER2	17	17
Equilibrated juvenile survival	N/A (A1 and A2)	67	57
	EJUV1- 0.85	17	18
	EJUV2 - 0.96	17	25
Transition juvenile survival	N/A (A1 and A2)	67	57
	TJUVa (1-2 Yrs)	17	26
	TJUVb (2-8 Yrs)	17	17
Transportation "Model"	0 (CRiSP)	50	61
	1 (FLUSH 71-89)	40	32
	2 (FLUSH, 0.83 * 1)	10	7
Prospective Model	Alpha	50	32
	Delta	50	68
Extra Mortality/Future Climate	BKD/Autoregressive	20	4.2
	BKD/Cyclical	20	9
	Hydro/Autoregressive	20	36.9
	Hydro/Cyclical	20	35.3
	Regime Shift/Cyclical	20	14.6
Habitat effects	0 (no improvement)	50	53
	B (max. protection)	50	47

*** Note that the these probabilities are summed across actions, possibly leading to an underestimate of the importance of some hypotheses.**

REVIEW DRAFT

Table 2. Testable (in principle) and “Important” hypotheses..

Uncertainty/Hypothesis	Hypothesis	Testable (in principle) with retrospective data through BY90?	Testable (in principle) with retrospective data through BY93?	Important (from Table 1)
Passage Model	PMOD1 – CRiSP	Y	Y	Y
	PMOD2 – FLUSH			
Fish Guidance Efficiency	ESBS > STS	N	Yes, for recent data	Y
	ESBS + STS		Yes, for recent data	
Turbine/Bypass Survival	TURB1	Y	Y	
	TURB4			
	TURB5			
	TURB6			
Predator Removal	PREM1 (no res. mortality reduction)	N	Yes, for recent data	Y
	PREM3 (25% reduction)	N	Yes, for recent data	
Pre-Removal Period	N/A (A1 and A2)	N	N	N
	PRER1			
	PRER2			
Equilibrated juvenile survival	N/A (A1 and A2)	N	N	N
	EJUV1- 0.85			
	EJUV2 - 0.96			
Transition juvenile survival	N/A (A1 and A2)	N	N	N
	TJUVa (1-2 Yrs)			
	TJUVb (2-8 Yrs)			
Transportation "Model"	0 (CRiSP)	Y	Y	N
	1 (FLUSH 71-89)			
	2 (FLUSH, $0.83 * 1$)			
Prospective Model	Alpha	Y	Y	Y
	Delta	Y	Y	
Extra Mortality/Future Climate	BKD/Autoregressive	N	N	Y
	BKD/Cyclical	N	N	
	Hydro/Autoregressive	Y (Hydro), N(Autoregressive)	Y (Hydro), N(Autoregressive)	
	Hydro/Cyclical	Y (Hydro), N(Autoregressive)	Y (Hydro), N(Autoregressive)	
	Regime Shift/Cyclical	Y (Regime Shift), N(Cyclical)	Y (Regime Shift), N(Cyclical)	
Habitat effects	0 (no improvement)	N	N	N
	B (max. protection)			

Table 3. BSM Models Testable with Retrospective Data

Life Cycle Model	Passage Model	Extra Mortality	Generic Model Equations (From Appendix A, sections 2 and 3)	Specific Equation for Hypothesis (by inference)
Alpha	CRiSP	Hydro	$\ln(R(i)) = (1+p)\ln(S(i)) + a(i) - b(i)S(i) - M + \alpha + \epsilon(i)$ $\alpha = \alpha(n) - \text{mean}(\alpha(n)) - \ln(DP + 1 - P) + \text{mean}(\ln(DP + 1 - P)) + \text{STEP}$ $\alpha(n) = C1/(F(n) - \text{mean}(F(n))) + C2(E(n)/F(n) - \text{mean}(E/F))$	$1 - \exp(-\text{STEP}) \propto 1 - V(n)$, so $\text{STEP} \propto -\ln(V(n))$. Note that proportion may not be constant from year to year.
	FLUSH	Hydro	As above	As above
	CRiSP	Regime Shift	As above	$\text{STEP}(75-90) \neq 0$, $\text{STEP}(52-74) = 0$
	FLUSH	Regime Shift	As above	As above
Delta	CRiSP	Hydro	$\ln(R(i)) = (1+p)\ln(S(i)) + a(i) - b(i)S(i) - m + \delta + \epsilon(i)$ $m = M + \Delta m$ $\omega = V(n)[P + \Phi(1-P)]$ $D = \Phi[V(n)/V(t)]$ $= \exp(-m)/\lambda(n)$ $\lambda(n) = \exp(-m - \ln(\omega))$ $\Phi, D, P, V(n)$ and $V(t)$ from passage model.	$1 - \lambda(n) \propto 1 - V(n)$, so $\lambda(n) \propto V(n)$. Note that proportion may differ from year to year.
	FLUSH	Hydro	As above	As above
	CRiSP	Regime Shift	As above	$\lambda(n) \propto \text{STEP}$ (0 for 52-74, 1 for 75-90)
	FLUSH	Regime Shift	As above	As above

Table 4. Spawner and Recruit Data to be used for “out-of-sample” tests.

Brood Year	Used in Retrospective Estimation?	Spawner Returns used to initialize BSM projections?	Spawning abundance estimates available now?	“Predict” spawners w/ BSM?	“Predict” recruits w/ BSM, with 98 returns?
1990	Yes	No	Yes	No	
1991	No	Yes	Yes	No	Yes, all ages
1992	No	Yes	Yes	No	Yes, all ages
1993	No	Yes	Yes	No	Yes, all ages
1994	No	Yes	Yes	No	Yes, through 2-ocean
1995	No	Yes	Yes	No	Yes, through 1-ocean
1996	No	No	Red Counts done but no spawner estimates yet	Yes	No
1997	No	No	Red Counts done but no spawner estimates yet	Yes	No
1998	No	No	Red Counts available late 1998	Yes	No

REVIEW DRAFT

Table 5. Mid-Columbia Spawner-Recruit Data

	<i>Wenatchee</i>			<i>Entiat</i>			<i>Methow</i>		
Brood	Spawners	Recruits	R/S	Spawners	Recruits	R/S	Spawners	Recruits	R/S
1955				388	1,184	3.06			
1956				223	1,135	5.09			
1957				409	961	2.35			
1958	1,377	9,451	6.86	142	1,451	10.21			
1959	1,102	10,504	9.53	296	2,808	9.47			
1960	2,225	21,910	9.85	316	5,162	16.31	2,006	19,466	9.70
1961	1,496	24,891	16.64	127	3,013	23.70	616	14,641	23.75
1962	2,733	28,726	10.51	315	3,930	12.48	2,472	17,720	7.17
1963	1,202	16,361	13.62	269	2,943	10.95	1,245	10,033	8.06
1964	2,758	15,476	5.61	1,096	2,281	2.08	3,845	7,246	1.88
1965	2,581	12,807	4.96	232	1,507	6.48	1,115	7,336	6.58
1966	5,930	7,922	1.34	831	1,012	1.22	4,280	7,766	1.81
1967	3,423	11,475	3.35	648	1,478	2.28	2,163	9,681	4.48
1968	4,239	17,144	4.04	685	1,683	2.46	1,707	10,743	6.30
1969	3,762	15,765	4.19	391	2,306	5.90	1,323	10,153	7.67
1970	2,565	15,680	6.11	182	2,576	14.12	1,525	12,492	8.19
1971	1,366	11,778	8.62	348	1,525	4.39	1,258	7,598	6.04
1972	2,734	7,773	2.84	182	1,077	5.94	1,569	8,321	5.30
1973	5,263	12,134	2.31	636	3,249	5.11	2,152	19,349	8.99
1974	2,392	8,322	3.48	267	3,537	13.22	1,163	14,317	12.31
1975	4,198	4,039	0.96	458	912	1.99	1,987	2,654	1.34
1976	2,069	5,284	2.55	81	1,084	13.39	390	2,013	5.15
1977	2,702	6,840	2.53	501	1,031	2.06	1,841	1,992	1.08
1978	3,943	5,312	1.35	1,009	783	0.78	2,541	1,296	0.51
1979	1,336	2,696	2.02	233	527	2.26	462	1,031	2.23
1980	1,742	4,725	2.71	295	400	1.35	348	2,699	7.75
1981	1,629	5,280	3.24	285	665	2.33	442	2,056	4.65
1982	1,822	3,898	2.14	322	691	2.15	528	5,120	9.71
1983	3,214	2,318	0.72	324	554	1.71	818	4,277	5.23
1984	2,369	1,775	0.75	250	117	0.47	868	3,986	4.59
1985	4,472	2,878	0.64	351	392	1.12	1,204	3,534	2.93
1986	2,713	1,553	0.57	321	351	1.09	891	2,099	2.36
1987	2,042	936	0.46	194	142	0.73	1,449	1,711	1.18
1988	1,875	1,254	0.67	201	382	1.89	1,588	4,742	2.99
1989	1,396	835	0.60	112	292	2.60	1,086	1,780	1.64
1990	1,021	42	0.04	254	99	0.39	1,089	159	0.15
1991	604			93			481		
1992	1,143			129			1,598		
1993	1,294			311			1,344		
1994	303			73			276		
1995	60			15			21		

Figure 1. Mid-Columbia Recruits per Spawner (Data from Table 5)